One way cuts in oriented graphs

Simon Griffiths

February 2, 2008

Abstract

A *one way cut* is a set of edges $E(A,B)$ from A to B for subsets $A,B \subset V$ for which $e(B,A)=0$. We write $\text{cut}(G)$ for the maximum size of a one way cut in G. Amini, Griffiths and Huc proved that for a random orientation G of $G(n,p)$ we have $\text{cut}(G)=\Theta(n)$. We prove that if G has no isolated vertices then $\text{cut}(G) \geq n/8\log n$. We then demonstrate that up to multiplication by constant this result is best possible, this is achieved by constructing a random oriented graph with a large range of degrees which with high probability has no isolated vertices and has $\text{cut}(G) \leq Kn/\log n$ for K a large constant. Similar results are proved for 'biased' cuts.

1 Introduction

For every simple graph G one can find disjoint subsets A,B of the vertex set such that $E(A,B)$ contains at least half of the edges of G. This follows immediately from the fact that $\mathbb{E}(e(A,B)) \geq e(G)/2$ for a random partition of the vertex set. In the context of oriented graphs (ie. digraphs without loops, digons or multiple edges) it is easy to show using the same probabilistic argument that there are disjoint subsets A,B for which $e(A,B) \geq e(G)/4$, where $e(A,B)$ now denotes the number of oriented edges from A to B.

We study a different question. A *one way cut* in an oriented graph G is a set of edges $E(A,B)$ from a set $A \subset V$ to a set $B \subset V$ for which $e(B,A)=0$. For an oriented graph G we consider the size of the largest one way cut,

$$\text{cut}(G) = \max\{e(A,B) : A,B \subset V \text{ with } e(B,A)=0\}$$

We shall also study, for $\nu \in (0,1)$, the largest value of $e(A,B)$ for which $e(B,A)/e(A,B) < \nu$, this is the largest cut with a strong 'bias' in one direction. We define,

$$\text{cut}_\nu(G) = \max\{e(A,B) : A,B \subset V \text{ with } e(B,A) < \nu e(A,B)\}$$

We begin by making some simple observations. From the definitions it follows immediately that for $0<\nu<\mu<1$ we have $\text{cut}(G) \leq \text{cut}_\nu(G) \leq \text{cut}_\mu(G)$. Another simple observation is that $\text{cut}(G) \geq \Delta^+(G)$, simply consider taking $A=\{v\}$ and $B=\Gamma^+(v)$ for some vertex v with out degree $\Delta^+(v)$. Similarly $\text{cut}(G) \geq \Delta^-(G)$.

We now recall some results of Amini, Griffiths and Huc, which make it clear that the study of one way cuts and of biased cuts is very different from the study of cuts which are not required to be biased. In particular it is not generally possible to find cuts containing a positive proportion of the edges of the graph.

Proposition 1.1. *[1] For each $\nu \in (0,1)$ there is a constant K_ν such that for any simple graph H, the oriented graph G obtained by orienting the edges of H at random satisfies $\mathrm{cut}_\nu(G) \leq K_\nu n$ with high probability.*

Proof. For $K \in \mathbb{R}$ let $W(K) = \{(A,B) : e_H(A,B) \geq Kn\}$. For $(A,B) \in W(K)$ we have by Chernoff's inequality that

$$\mathbb{P}(e(B,A) < \nu e(A,B)) \leq \exp(-c(\nu)e_H(A,B)) \leq \exp(-c(\nu)Kn)$$

Setting $K_\nu = 2/c(\nu)$ and noting $|W(K_\nu)| \leq 4^n$ we have

$$\mathbb{P}(e(B,A) < \nu e(A,B) \text{ for some } (A,B) \in W(K_\nu)) \leq 4^n \exp(-2n)$$

so that with high probability we have $e(B,A) \geq \nu e(A,B)$ for every pair A, B for which $e_H(A,B) \geq K_\nu n$. Hence with high probability $\mathrm{cut}_\nu(G) \leq K_\nu n$. □

Of course this also implies the existence of a constant K for which $\mathrm{cut}(G) \leq Kn$ with high probability. A simple calculation shows that any $K > 2$ works. We can also put lower bounds on $\mathrm{cut}(G)$, in particular,

Proposition 1.2. *[1] Let G be a regular oriented graph (ie. for some $d \in \mathbb{N}$ we have $d^+(v) = d^-(v) = d$ for all $v \in V$) then $\mathrm{cut}(G) \geq n/4$*

Proof. For a set $A \subset V$ we define $B(A) = \{v \in V : e(\{v\}, A) = 0\}$, it follows that $e(B(A), A) = 0$. We show the existence of a set $A \subset V$ for which $e(A, B(A)) \geq n/4$. Let A be a random subset of V with each v being an element of A independently with probability p. Then,

$$\mathbb{E}(e(A,B(A))) = \sum_{xy \in E(G)} \mathbb{P}(x \in A, y \in B(A)) = \sum_{E(G)} \mathbb{P}(x \in A)\mathbb{P}(A \cap \Gamma^+(y) = \phi) = p(1-p)^d e \geq p(1-dp)e$$

Setting $p = 1/2d$ we obtain $\mathbb{E}(e(A,B(A))) \geq e/4d = n/4$, and so there exists a set $A \subset V$ for which $e(A,B(A)) \geq n/4$. □

The same proof shows that $\mathrm{cut}(G) \geq n/8$ for oriented graphs with the property of being almost regular, ie. that there is a value of d such that $d \leq d^+(v), d^-(v) \leq 2d$ for all $v \in V$. Let $G(n,p)$ the random (simple) graph in which edges are included independently with probability p. For $p = \omega(\log n/n)$ a random orientation G of $G(n,p)$ is, with high probability, almost regular, and so with high probability $\mathrm{cut}(G) \geq n/8$.

Proposition 1.2 works for regular graphs of all degrees. Therefore the number of edges in the oriented graph does not seem to be an important parameter in problems of this type. What seems more important is that the condition of regularity ensures that G does not have isolated vertices. The author believed for some time that it should be possible to prove a version of

Proposition 1.2 for oriented graphs without isolated vertices, ie. show for some constant $c > 0$ that $\text{cut}(G) \geq cn$ for all oriented graphs without isolated vertices. Certainly it is possible to show $\text{cut}(G) \geq n/8\lceil \log n \rceil$, here and throughout the article $\log n$ is the logarithm base 2.

Proposition 1.3. *Let G be an oriented graph without isolated vertices then* $\text{cut}(G) \geq n/8\lceil \log n \rceil$

Remark. To see that a condition on isolated vertices is more useful than a condition on the number of edges, let G be a random orientation of a clique on $n^{0.9}$ vertices then G has quite a lot of edges while by Proposition 1.1 (or really, by going through the proof of Proposition 1.1 with $K = 3$) we have that $\text{cut}(G) \leq 3n^{0.9}$.

Proof. We partition the vertex set V into $V^+ = \{x \in V : d^+(x) \geq d^-(x)\}$ and $V^- = \{x \in V : d^-(x) > d^+(x)\}$. We then partition each of the sets V^+ and V^- as follows we define $V_i^+ = \{x \in V^+ : n/2^i < d^+(x) \leq n/2^{i-1}\}$ and $V_i^- = \{x \in V^- : n/2^i < d^-(x) \leq n/2^{i-1}\}$ for $i = 1, ..., \lceil \log n \rceil$. By the pigeon hole principle one of these $2\lceil \log n \rceil$ sets has size at least $n/2\lceil \log n \rceil$, we may assume it is V_i^- for some i (otherwise proceed with the same argument but with A and B interchanged) we then pick a random subset A of V with each vertex included in A independently with probability p and define $B(A) = \{v \in V_i^- : e(\{v\}, A) = 0\}$ then $e(B(A), A) = 0$ and,

$$\mathbb{E}(e(A, B(A)) = \sum_{y \in V_i^-} \sum_{xy \in E} \mathbb{P}(x \in A)\mathbb{P}(y \in B(A)) = \sum_{y \in V_i^-} d^-(y)p(1-p)^{d^+(y)} \geq \frac{n}{2\lceil \log n \rceil} \frac{np}{2^i} \text{cut}(G) \geq \left(1 - \frac{np}{2^{i-1}}\right)$$

setting $p = 2^{i-2}/n$ we obtain $\mathbb{E}\big(e(A, B(A))\big) \geq n/8\lceil \log n \rceil$, and so there exists $A \subset V$ with $e(A, B(A)) \geq n/8$. $\square$

The main result is that remarkably, up to a change of the constant, the result of Proposition 1.3 is best possible.

Theorem 1.4. *There exists a constant $K \in \mathbb{R}$ such that for each $n \in \mathbb{N}$, there exists an oriented graph G on n vertices with no isolated vertices and with* $\text{cut}(G) \leq Kn/\log n$

For an oriented graph G and a subset $A \subset V$ we define $B(A) = \{v : e(\{v\}, A) = 0\}$. For a fixed A it is clear that $B = B(A)$ maximises $e(A, B)$ amongst sets B for which $e(B, A) = 0$, since $B(A)$ contains all sets B for which $e(B, A) = 0$. So that our aim is to find an oriented graph G without isolated vertices for which $e(A, B(A)) \leq Kn/\log n$ for all $A \subset V$. It is clear from Proposition 1.2 that we cannot take G to be regular. In fact if G is close to regular in the sense that all in and out degrees are of the same order (ie. there exists k such that $\Delta^+/\delta^+ \leq k$ and $\Delta^-/\delta^- \leq k$) then by adapting the proof of Proposition 1.3 it can be shown that $\text{cut}(G) > cn$ for a constant $c = c(k) > 0$.

Hence our example must have a large range of degrees. It is also clear that our example G should be random-like. We meet these two criteria by defining a random graph with a large range of degrees. We do this in Section 2. We then demonstrate that $\text{cut}(G) \leq Kn/\log n$, we do this by first demonstrating that for no set A do we have $B(A)$ very much larger than we expect, we then obtain the required result using a large deviation inequality.

Furthermore our example will also have no large biased cuts, specifically,

Theorem 1.5. *For $\nu \in (0,1)$ there exists a constant K_ν such that for each $n \in \mathbb{N}$, there exists an oriented graph G on n vertices with no isolated vertices and with $\mathrm{cut}_\nu(G) \leq K_\nu n/\log n$*

The proof of this result, which appears in Section 3 requires a little more effort than the proof of Theorem 1.4, but is not a fundamentally different proof.

2 Proof of Theorem 1.4

We are about to define the oriented graph G required to demonstrate the Theorem. In fact we choose G to be bipartite, this is not an essential property of G, but it seems to make the rest of the proof easier to visualise. Set $K = 100001$. The Theorem now trivially holds for all $n < K/\log K$ and so for all $n \leq 2^{12}$. Note that if G is an example on n vertices then adding a leaf (ie. a vertex joined to G by a single edge) gives an example G' on $n+1$ vertices with $\mathrm{cut}(G') \leq \mathrm{cut}(G)+1$, so that it suffices to prove the theorem for even n with $K = 100000$. For even $n \geq 2^{12}$ we define G as follows.

Let l be the largest even integer with $l \leq \log(n)$, so that $l/2$ is an integer. We define disjoint sets $U_1, ..., U_{l/2}$ of cardinality approximately n/l, formally we write $n/2$ as $ql/2 + r$ where $r \in [0, ..., l/2)$, and define the cardinalities of the sets U_i by,

$$|U_i| = q+1 \text{ for } i = 1, ..., r \quad \text{ and } \quad |U_i| = q \text{ for } i = r+1, ..., l/2 \tag{1}$$

Similarly we define $V_1, ..., V_{l/2}$ to be disjoint sets with cardinalities,

$$|V_j| = q+1 \text{ for } j = 1, ..., r \quad \text{ and } \quad |V_j| = q \text{ for } j = r+1, ..., l/2 \tag{2}$$

It is inconvenient that we cannot just say that we cannot just say that each set U_i, V_j has size n/l, but we do have (from the assumption $n \geq 2^{12}$) that

$$\frac{n}{l\alpha} \leq |U_i|, |V_j| \leq \frac{\alpha n}{l} \quad \text{ for all } i, j \text{ where } \alpha = 1 + \frac{1}{2^8} \tag{3}$$

The vertex set of G is $U \cup V$ where,

$$U = \bigcup_{i=1}^{l/2} U_i \quad \text{ and } \quad V = \bigcup_{j=1}^{l/2} V_j \tag{4}$$

We now define the edge set $E(G)$ randomly as follows, for $u \in U_i$ and $v \in V_j$ we include the edge uv with probability $1/2^{i+j}$ and vu with probability $1/2^{i+j}$, we demand that these events are disjoint to ensure that G is an oriented graph. However this process is done independently for each pair $(u,v) \in U \times V$. Equivalently, a simple graph is formed at random with edge uv included with probability $1/2^{i+j-1}$ (where $u \in U_i, v \in V_j$) and independently of other pairs, and is then oriented at random.

Remark. It is unfortunate that we do not have complete independence. In our arguments we will consider first the set of edges going in one direction (say from V to U) and then the set

of edges going in the other direction, we observe that for a set E' of edges from V to U and a pair uv (say $u \in U_i$, $v \in V_j$) for which $vu \notin E'$ we have that

$$\frac{1}{2^{i+j}} \leq \mathbb{P}(uv \in E(G)|E(V,U) = E') \leq \frac{1}{2^{i+j-1}} \tag{5}$$

. We do not need to worry about the details yet, but we will wish to bound $e(A, B(A))$. Starting with a set $A \subset U$, we consider the 'size' of $B(A) \subset V$, this relies on the set $E(V,U)$. We then know that for $u \in A \cap U_i$ and $v \in B(A) \cap V_j$ we have $1/2^{i+j} \leq \mathbb{P}(uv \in E(G)) \leq 1/2^{i+j-1}$. This enables us to bound $e(A, B(A))$.

We show that with high probability G has no isolated vertices and has $\mathrm{cut}(G) \leq Kn/l$, so that the there exists a choice of G with these two properties, and the Theorem is proved. We first show that with high probability G has no isolated vertices.

Lemma 2.1. $\mathbb{P}(G$ *has an isolated vertex*$) \leq 2\exp(l - n^{1/2}/2l)$. *In particular, with high probability, G has no isolated vertex*

Proof. Let $u \in U_i$ for some $i = 1, ..., l/2$, for each $v \in V_1$ the probability that there is an edge between u and v (in either direction) is $1/2^i$. It follows that,

$$\mathbb{P}(e(\{u\}, V_1) = 0) = \left(1 - \frac{1}{2^i}\right)^{|V_1|} \leq \left(1 - \frac{1}{2^i}\right)^{n/l} \leq \exp\left(\frac{-n}{l2^i}\right) \leq \exp\left(\frac{-n}{l2^{l/2}}\right) \leq \exp\left(\frac{-n^{1/2}}{2l}\right)$$

and so

$$\mathbb{P}(\text{there is an isolated vertex in } U) \leq \sum_{u \in U} \mathbb{P}(e(\{u\}, V_1) = 0) = \frac{n}{2}\exp\left(\frac{-n^{1/2}}{2l}\right) \leq \exp\left(l - \frac{n^{1/2}}{2l}\right)$$

using a similar argument the probability that there is an isolated vertex in V may also be bounded by $\exp(l - n^{1/2}/2l)$. □

Let $\mathrm{cut}(G; U, V) = \max\{e(A, B) : A \subset U, B \subset V \text{ with } e(B, A) = 0\}$, then it is immediate that $\mathrm{cut}(G) \leq \mathrm{cut}(G; U, V) + \mathrm{cut}(G; V, U)$ so it suffices to show that with high probability each of $\mathrm{cut}(G; U, V)$ and $\mathrm{cut}(G; V, U)$ is at most $Kn/2l$. By symmetry it suffices to show that $\mathrm{cut}(G; U, V) \leq Kn/2l$ with high probability, we shall not carry this factor of 2 with us throughout the proof, we shall show $\mathrm{cut}(G; U, V) \leq Kn/l$ with high probability for $K \geq 50000$. For a set $A \subset U$ the maximum value of $e(A, B)$ for $B \subset V$ with $e(B, A) = 0$ is obtained by the set $B(A) = \{v \in V : e(\{v\}, A) = 0\}$. So that the event $\mathrm{cut}(G; U, V) > Kn/l$ is contained in the event $\bigcup_{A \subset U} F(A)$ where $F(A)$ is the event that $e(A, B(A)) > Kn/l$, hence to prove Theorem 1.4 we just need to prove,

Lemma 2.2. $\mathbb{P}(\bigcup_{A \subset U} F(A)) \leq 2l^2 n^{2l} \exp(-n/l)$

For a fixed A what can we say about the probability of the event $F(A)$, that $e(A, B(A)) > Kn/l$? We'll first show $\mathbb{E}(e(A, B(A)) \leq 7n/l$.

Notation. For a set $A \subset U$, let $A_i = A \cap U_i$ and $a_i = |A_i|$ for $i = 1, ..., l/2$. We then set $s(A) = \sum_i a_i/2^i$, this value is measure of the 'size' of the set A in the sense that the expected number of edges from A to V (or to a given V_j or a vertex v of V) is proportional to $s(A)$. Similarly for $B \subset V$ we let $B_j = B \cap V_j$ and $b_j = |B_j|$ for $j = 1, ..., l/2$. We set $t(B) = \sum_j b_j/2^j$, and then $t(B)$ is a measure of the size of B in the same way that $s(A)$ is a measure of the size of A. In fact we can use these values together and we discover that the expected number of edges from A to B is proportional to $s(A)t(B)$.

Lemma 2.3. *Let $A \subset U$ then $\mathbb{E}(e(A, B(A))) \leq 7n/l$*

Proof. Let $B = B(A)$ and let $s = s(A) = \sum_i a_i/2^i$. By linearity of expectation $\mathbb{E}(e(A, B)) = \sum_j \mathbb{E}(e(A, B_j))$. For each j we have that $\mathbb{E}(e(A, B_j)) = \sum_{v \in V_j} \mathbb{P}(v \in B_j)\mathbb{E}(e(A, \{v\})|v \in B_j)$, for each $v \in V_j$ we have,

$$\mathbb{P}(v \in B_j) = \mathbb{P}(e(\{v\}, A) = 0) = \prod_{i=1}^{l/2} \mathbb{P}(e(\{v\}, A_i) = 0) = \prod_i \left(1 - \frac{1}{2^{i+j}}\right)^{a_i} \tag{6}$$

Using the inequality $(1 - x)^n \leq \exp(-xn)$ we obtain,

$$\mathbb{P}(v \in B_j) \leq \prod_i \exp\left(\frac{-a_i}{2^{i+j}}\right) = \exp\left(-\sum_i \frac{a_i}{2^{i+j}}\right) = \exp\left(\frac{-s}{2^j}\right) \tag{7}$$

now for each pair $(u, v) \in A_i \times B_j$ we have (by the remark above) that the probability uv is an edge is at most 2^{-i-j+1}, so that for $v \in B_j$ we have

$$\mathbb{E}(e(A, \{v\})) = \sum_i \mathbb{E}(e(A_i, \{v\})) \leq \sum_i \mathbb{E}(\mathrm{Bin}(a_i, 2^{-i-j+1})) = \frac{1}{2^{j-1}} \sum_i \frac{a_i}{2^i} = \frac{s}{2^{j-1}}$$

summing over vertices $v \in B_j$ we obtain

$$\mathbb{E}(e(A, B)) = \sum_j \mathbb{E}(e(A, B_j)) \leq 2 \sum_j \mathbb{E}(|B_j|)\frac{s}{2^j} \leq \frac{2\alpha n}{l} \sum_j \frac{s}{2^j} \exp\left(\frac{-s}{2^j}\right)$$

The sum simply samples the function $x \exp(-x)$ at the values $s/2^j$, writing j_0 for the choice of j such that $2^j \leq s < 2^{j+1}$ it is not difficult to bound the sum by a constant,

$$\sum_{j=1}^{l/2} \frac{s}{2^j} \exp(\frac{-s}{2^j}) \leq \sum_{-\infty}^{j_0} \exp(\frac{-s}{2^j}) + \sum_{j_0+1}^{\infty} \frac{s}{2^j} \leq 1 + 2 = 3$$

so that $\mathbb{E}(e(A, B)) \leq 6\alpha n/l < 7n/l$ □

One might now hope to prove Lemma 2.2 by using a concentration inequality to bound $\mathbb{P}(F(A))$ for each $A \subset U$ and using the bound $\mathbb{P}(\bigcup_{A \subset U} F(A)) \leq \sum_{A \subset U} \mathbb{P}(F(A))$. Unfortunately this approach cannot be successful, the event $F(A)$ that $e(B, A) > Kn/l$ is indeed very unlikely, an application of the Chernoff inequality shows that

$$\mathbb{P}(e(A, B(A)) > Kn/l) \leq \exp(-c(K)n/l) \tag{8}$$

where $c(K)$ is a constant dependent on K such that $c(K) \to \infty$ as $K \to \infty$. However there are $2^{n/2}$ choices of $A \subset U$, so that one would only obtain $\mathbb{P}(\bigcup_{A\subset U} F(A)) \leq 2^{n/2} \exp(-c(K)n/l)$, which is trivial for large n. Furthermore, it is not possible for a general $A \subset U$ to put a significantly stronger upper bound on $\mathbb{P}(e(A, B(A)) > Kn/l)$, it is certainly not in general true that $\mathbb{P}(e(A, B(A)) > Kn/l) < 2^{-n/2}$.

The above approach failed because we had to consider all $2^{n/2}$ subsets $A \subset U$ separately, when in fact there are many sets $A \subset U$ which are essentially very similar from the point of view of this problem. We shall prove results for a collection of special sets A and show that this is sufficient to give the result for all $A \subset U$. To capture information about many sets $A \subset U$ simultaneously we shall use the following monotonicity result.

Lemma 2.4. *For subsets $A^0 \subset A \subset A^1$ of U we have $e(A, B(A)) \leq e(A^1, B(A^0))$. So that the event $F(A)$ that $e(A, B(A)) > Kn/l$ is contained in the event $F(A^0, A^1)$ that $e(A^1, B(A^0)) > Kn/l$.*

Proof. Since $B(A)$ is monotone decreasing in A we have that $B(A) \subset B(A^0)$. Now since $e(A, B)$ is monotone increasing in A and B we have $e(A, B(A)) \leq e(A, B(A^0)) \leq e(A^1, B(A^0))$ $\square$

We now define some collections of special sets which play the role of A^0. For each $i_0 \in \{1, ..., l\}$ we define $i_0' = \min\{i_0, l/2\}$, the interval $I_{i_0} = [2n/2^{i_0}l, 8n/2^{i_0}l)$ and the collection,

$$\mathcal{A}(i_0) = \{A \subset \bigcup_{i=1}^{i_0'} U_i : s(A) \in I_{i_0}\} \tag{9}$$

We also define a partition of $\mathcal{A}(i_0)$. For a sequence $a_1, ..., a_{i_0'}$ we let $s(a_1, ..., a_{i_0'}) = \sum_i a_i/2^i$ and write $S(i_0)$ for the set of sequences $(a_1, ..., a_{i_0'})$ with $a_i \leq |U_i|$ for each i and with $s(a_1, ..., a_{i_0'}) \in I_{i_0}$. For each $(a_1, ..., a_{i_0'}) \in S(i_0)$ we define,

$$\mathcal{A}(i_0; a_1, ..., a_{i_0'}) = \{A \in \bigcup_{i=1}^{i_0'} U_i : |A_i| = a_i \text{ for all } i\} \tag{10}$$

Of course we may now partition $\mathcal{A}(i_0)$ as follows

$$\mathcal{A}(i_0) = \bigcup_{(a_1,...,a_{i_0'})\in S(i_0)} \mathcal{A}(i_0; a_1, ..., a_{i_0}) \tag{11}$$

For each $i_0 \in \{1, ..., l\}$ and each $A \in \mathcal{A}(i_0)$ let $F(i_0; A)$ to be the event that

$$e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > Kn/l$$

It follows from Lemma 2.4 that the event $F(i_0; A)$ contains all the events $F(A')$ with $A \subset A' \subset A \cup \bigcup_{i_0+1}^{l/2} U_i$. For this reason it is useful to note the following.

Lemma 2.5. *For each non-empty subset $A' \subset U$ there is a choice of $i_0 \in \{1,...,l\}$ and $A \in \mathcal{A}(i_0)$ with $A \subset A' \subset A \cup \bigcup_{i_0+1}^{l/2} U_i$*

Proof. Let i_0 be chosen such that $s(A') \in [4n/2^{i_0}l, 8n/2^{i_0}l)$, and note that $\sum_{i>i_0} |A'_i|/2^i \leq \frac{\alpha n}{l}\sum_{i>i_0} 2^{-i} = \frac{\alpha n}{2^{i_0}l}$. Let $A = A' \cap \bigcup_{i=1}^{i'_0} U_i$. Now,

$$s(A) = \sum_{i=1}^{i'_0} \frac{|A_i|}{2^i} = \sum_{i=1}^{i'_0} \frac{|A'_i|}{2^i} = s(A') - \sum_{i>i_0} \frac{|A'_i|}{2^i} \geq \frac{2n}{2^{i_0}l} \tag{12}$$

and $s(A) \leq s(A') \leq 8n/2^{i_0}l$, so that $s(A) \in I(i_0)$. We now have that $A \in \mathcal{A}(i_0)$ and $A \subset A' \subset A \cup \bigcup_{i=i_0+1}^{l/2} U_i$. $\square$

From Lemma 2.4 and Lemma 2.5 it follows that the event $\bigcup_{A \subset U} F(A)$ is contained in the event $\bigcup_{i_0=1}^{l} \bigcup_{A \in \mathcal{A}(i_0)} F(i_0; A)$. Using our partition of each $\mathcal{A}(i_0)$ we may write this as

$$\bigcup_{i_0=1}^{l} \bigcup_{(a_1,...,a_{i'_0}) \in S(i_0)} \bigcup_{A \in \mathcal{A}(i_0; a_1,...,a_{i'_0})} F(i_0; A) \tag{13}$$

alternatively writing $F(i_0; a_1, ..., a_{i'_0})$ for the union of $F(i_0; A)$ over sets $A \in \mathcal{A}(i_0; a_1, ..., a_{i'_0})$ this is,

$$\bigcup_{i_0=1}^{l} \bigcup_{(a_1,...,a_{i'_0}) \in S(i_0)} F(i_0; a_1, ..., a_{i'_0}) \tag{14}$$

How many events $F(i_0; a_1, ..., a_{i'_0})$ occur in this union? There are l choices of i_0 and for each i_0 there are at most $(\alpha n/l)^{i'_0} \leq n^l$ choices of the sequence $(a_1, ..., a_{i'_0})$. So there are at most ln^l events $F(i_0; a_1, ..., a_{i'_0})$ in the union. Hence to prove Lemma 2.2 it suffices to prove the following Lemma.

Lemma 2.6. *For each $i_0 \in \{1,...,l\}$ and $(a_1, ..., a_{i'_0}) \in S_{i_0}$ we have $\mathbb{P}(F(i_0; a_1, ..., a_{i'_0})) \leq 2ln^l \exp(-n/l)$.*

For $A \in \mathcal{A}(i_0; a_1, ..., a_{i'_0})$ let us express the event $F(i_0; A)$ that $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > Kn/l$ as a union of two events

$$F(i_0; A) = F_B(i_0; A) \cup (F(i_0; A) \setminus F_B(i_0; A)) \tag{15}$$

where $F_B(i_0; A)$ is the event that $t(B(A)) > 2^{i_0+10}$. The event $F(i_0; A)$ that $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > Kn/l$ depends on the (random) collection of edges from V to U as this collection of edges defines $B(A)$ and the (random) collection of edges from U to V as this determines $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$ for the set $B(A)$ that has been defined. We have split the event $F(i_0; A)$ so that we may consider these two incidences of randomness separately. $F_B(i_0; A)$ should be thought of as the event that $B(A)$ is much 'larger' than it should be. While $(F(i_0; A) \setminus F_B(i_0; A))$ is the event

that there are very many edges from $A \cup \bigcup_{i_0+1}^{l/2} U_i$ to $B(A)$ even though $B(A)$ is not so large. We obtain our results by showing that each of these events is very unlikely.

We define $F_B(i_0; a_1, ..., a_{i_0'})$ to be the event that $F_B(i_0; A)$ occurs for some $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ and express the event $F(i_0; a_1, ..., a_{i_0'})$ as the union,

$$F(i_0; a_1, ..., a_{i_0'}) = F_B(i_0; a_1, ..., a_{i_0'}) \cup (F(i_0; a_1, ..., a_{i_0'}) \setminus F_B(i_0; a_1, ..., a_{i_0'})) \tag{16}$$

We shall prove Lemma 2.6 by showing each of these events has probability at most $ln^l \exp(-n/l)$, we begin with $F_B(i_0; a_1, ..., a_{i_0'})$.

Lemma 2.7. *Let $i_0 \in \{1, ..., l\}$ and $(a_1, ..., a_{i_0'}) \in S(i_0)$ then $\mathbb{P}(F_B(i_0; a_1, ..., a_{i_0'})) \leq ln^l \exp(-n/l)$*

The event is a union over $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ of events $F_B(i_0; A)$. Let us first put an upper bound on $|\mathcal{A}(i_0; a_1, ..., a_{i_0'})|$ that will be used in the proof of Lemma 2.7.

Lemma 2.8. *Let $i_0 \in \{1, ..., l\}$ and $(a_1, ..., a_{i_0'}) \in S(i_0)$ then $|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \exp(165n/l)$*

Proof. Using the well known inequality $\binom{n}{k} \leq (en/k)^k$, where e is the base of the natural logarithm, we may bound $|\mathcal{A}(i_0; a_1, ..., a_{i_0'})|$ as follows,

$$|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \prod_{i=1}^{i_0'} \binom{\lceil n/l \rceil}{a_i} \leq \prod_{i=1}^{i_0'} \left(\frac{e\lceil n/l \rceil}{a_i}\right)^{a_i} \leq \prod_{i=1}^{i_0'} \left(\frac{e\alpha n}{la_i}\right)^{a_i} \leq \prod_{i=1}^{i_0'} \exp\left(\left(2+\log\left(\frac{n}{la_i}\right)\right)a_i\right) \tag{17}$$

so that

$$|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \exp\left(\sum_{i=1}^{i_0'} \left(2 + \log\left(\frac{n}{la_i}\right)\right)a_i\right) \tag{18}$$

We note that for each i we have $a_i/2^i \leq s(A) \leq 8n/l2^{i_0}$ so that $a_i \leq 8n/l2^{i_0-i}$ it follows that $\sum_{i=1}^{i_0'} a_i \leq (8n/l)\sum_{i=1}^{i_0'} 2^{i_0-i} \leq 16n/l$ so that $\exp(2\sum a_i) \leq \exp(32n/l)$ and our estimate on $|\mathcal{A}(i_0; a_1, ..., a_{i_0'})|$ can be simplified to

$$|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \exp\left(\frac{32n}{l} + \sum_{i=1}^{i_0'} a_i \log\left(\frac{n}{la_i}\right)\right) \tag{19}$$

we now bound $\sum_{i=1}^{i_0'} a_i \log(n/la_i)$ by writing it as the sum of two sums and then bounding each individually,

$$\sum_{i=1}^{i_0'} a_i \log\left(\frac{n}{la_i}\right) = \sum_{i=1}^{i_0'} a_i 2^{i_0-i+4} + \sum_{i=1}^{i_0'} a_i \left(\log\left(\frac{n}{la_i}\right) - 2^{i_0-i+4}\right) \tag{20}$$

The first sum is immediately bounded as follows,

$$\sum_{i=1}^{i_0'} a_i 2^{i_0-i+4} = 2^{i_0+4} s(A) \leq 2^{i_0+4} \frac{8n}{2^{i_0} l} = \frac{2^7 n}{l} \tag{21}$$

A term in the second sum makes a positive contribution only if $\log(n/la_i) > 2^{i_0-i+4}$. For this reason we define,

$$I_+ = \{i \leq i_0' : \log\left(\frac{n}{la_i}\right) > 2^{i_0-i+4}\} \tag{22}$$

For all $i \in I_+$ we have $\log(n/la_i) > 2^{i_0-i+4} \geq 2^4$, so that $n/la_i > 2^{2^4} = 2^{16}$ and so

$$a_i < \frac{n}{2^{16}l} \tag{23}$$

. We now partition I_+ . For $k \in \mathbb{N}$ we write $\log_k n$ for the kth iterated logarithm (base 2) of n. Let k_0 be the largest integer k for which $\log_k n \geq 2^8$. We let $I_0 = \{i \in I_+ : a_i \leq n/l(\log n)^2\}$, for $1 \leq k \leq k_0$ we let

$$I_k = \{i \in I_+ : n/l(\log_k n)^2 < a_i \leq n/l(\log_{k+1} n)^2\} \tag{24}$$

It is now clear (using (23)) that $I_0, ..., I_{k_0}$ is a partition of I_+. We bound the sum in (20) by bounding for each k the sum over $i \in I_k$. For $i \in I_0$ we have $a_i \leq n/l(\log n)^2$ and $\log(n/la_i) \leq \log n$ so that $a_i \log(n/la_i) \leq n/l \log n$, while $|I_0| \leq l/2 \leq \log n$ so that

$$\sum_{i \in I_0} a_i \log\left(\frac{n}{la_i}\right) \leq \frac{n}{l} \tag{25}$$

while for $1 \leq k \leq k_0$ for each $i \in I_k$ we have $a_i \leq n/l(\log_{k+1} n)^2$ and $\log(n/la_i) \leq \log((\log_k n)^2) = 2\log_{k+1} n$, so that,

$$a_i \log\left(\frac{n}{la_i}\right) \leq \frac{2n\log_{k+1} n}{l(\log_{k+1} n)^2} = \frac{2n}{l\log_{k+1} n} \tag{26}$$

while we may bound $|I_k|$ as follows, for $i \in I_k$ we have $a_i > n/l(\log_k n)^2$, and also (since $I_k \subset I_+$) we have that $\log n/la_i > 2^{i_0-i+4}$, putting these together we have that

$$2^{i_0-i+4} < \log\left(\frac{n}{la_i}\right) \leq \log((\log_k n)^2) = 2\log_{k+1} n \tag{27}$$

taking logs this implies $i_0 - i + 4 < \log_{k+2} n + \log 2$ and so $i > i_0 - \log_{k+2} n$. We also have that $i \leq i_0' \leq i_0$ so that for all $i \in I_k$ we have $i \in (i_0 - \log_{k+2} n, i_0]$ and so $|I_k| \leq \log_{k+2} n$. Putting this together with (26) we have that,

$$\sum_{i \in I_k} a_i \log\left(\frac{n}{la_i}\right) \leq \frac{2n\log_{k+2} n}{l\log_{k+1} n} \tag{28}$$

A bound for the sum in (20) can now be given,

$$\sum_{i \in I_+} a_i \log\left(\frac{n}{la_i}\right) = \sum_{k=0}^{k_0} a_i \log\left(\frac{n}{la_i}\right) \leq \frac{n}{l} + \frac{2n}{l}\sum_{k=1}^{k_0} \frac{\log_{k+2} n}{\log_{k+1} n} \tag{29}$$

Since $\log x/x < x^{-1/2}$ for all $x \geq 8$ we have for all $1 \leq k \leq k_0$ that $\log_{k+2} n/\log_{k+1} n < (\log_{k+1} n)^{-1/2}$. Let $x = \log_{k_0+1} n$ we have that $2^x = \log_{k_0} n$, $2^{2^x} = \log_{k_0-1} n$ and so on, so that $\log_{k_0-j} n$ is obtained by a tower of $j+1$ twos followed by x. It is immediately clear that for

all $j = 0, ..., k_0 - 1$ we have $\log_{k_0-j} n \geq 2^{jx}$ and so by the fact that the function $f(y) = y^{-1/2}$ is monotone decreasing we have

$$\sum_{k=1}^{k_0} \frac{\log_{k+2} n}{\log_{k+1} n} \leq \sum_{j=0}^{k_0-1} 2^{-jx/2} \tag{30}$$

this is a sum of a geometric progression with common ratio $2^{-x/2} < 1/2$ so that the value of the sum is at most twice the first term, so is at most 2. So that

$$\sum_{i \in I_+} a_i \log\left(\frac{n}{la_i}\right) = \sum_{k=0}^{k_0} \sum_{i \in I_k} a_i \log\left(\frac{n}{la_i}\right) \leq \frac{n}{l} + \frac{2n}{l} \sum_{k=1}^{k_0} \frac{\log_{k+2} n}{\log_{k+1} n} \leq \frac{5n}{l} \tag{31}$$

We now see that the sum in (20) is at most $2^7 n/l + 5n/l = 133n/l$. Substituting into (19) we obtain

$$|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \exp\left(\frac{32n}{l} + \frac{133n}{l}\right) = \exp\left(\frac{165n}{l}\right) \tag{32}$$

□

We now proceed to a proof of Lemma 2.7.

Proof of Lemma 2.7. The event $F_B(i_0; a_1, ..., a_{i_0'})$ is the union over sets $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ of the event $F_B(i_0; A)$. It is clear that the probability of the event $F_B(i_0; A)$ is the same for all sets $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$. Fixing $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ we have that

$$\mathbb{P}(F_B(i_0; a_1, ..., a_{i_0'})) = |\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \mathbb{P}(F_B(i_0; A)) \tag{33}$$

From Lemma 2.8 we have

$$|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \exp(165n/l) \tag{34}$$

Therefore our aim is to prove,

$$\mathbb{P}(F_B(i_0; A)) \leq ln^l \exp(-166n/l) \tag{35}$$

In doing this we shall partition $\{B : B \subset V\}$ as we have $\{A : A \subset U\}$. For each $j_0 \in \{1, ..., l\}$ we define $j_0' = \min\{j_0, l/2\}$, the interval $I_{j_0} = [2n/2^{j_0} l, 8n/2^{j_0} l)$ and the collection,

$$\mathcal{B}(j_0) = \{B \subset \bigcup_{j=1}^{j_0'} V_j : t(B) \in I_{j_0}\} \tag{36}$$

We now define a partition of $\mathcal{B}(j_0)$. For a sequence $b_1, ..., b_{j_0'}$ we let $t(b_1, ..., b_{j_0'}) = \sum_j b_j/2^j$ and write $T(j_0)$ for the set of sequences $(b_1, ..., b_{j_0'})$ with $b_j \leq |V_j|$ for each j and with $t(b_1, ..., b_{j_0'}) \in I_{j_0}$. For each $(b_1, ..., b_{j_0'}) \in T(j_0)$ we define,

$$\mathcal{B}(j_0; b_1, ..., b_{j_0'}) = \{B \in \bigcup_{j=1}^{j_0'} V_j : |B_j| = b_j \text{ for all } j\} \tag{37}$$

Of course we may now partition $\mathcal{B}(j_0)$ as follows

$$\mathcal{B}(j_0) = \bigcup_{(b_1,...,b_{j_0'})\in T(j_0)} \mathcal{B}(j_0; b_1, ..., b_{j_0'}) \tag{38}$$

For any set $B' \subset V$ there is a choice of j_0 such that setting $B = B' \cap \bigcup_{j=1}^{j_0'} V_j$ gives a set B with the properties that $B \subset B'$, that $t(B) \geq t(B')/2$ and $B \in \mathcal{B}(j_0)$. For a proof, simply choose j_0 such that $t(B') \in [4n/2^{j_0}l, 8n/2^{j_0}l)$ and note that $\sum_{j>j_0} |B_j'|/2^j \leq \frac{\alpha n}{l} \sum_{j>j_0} 2^{-j} \leq \frac{\alpha n}{2^{j_0}l} < t(B')/2$. Note also that $t(B) = t(B') - \sum_{j>j_0} |B_j'|/2^j$ and so $t(B) \geq t(B')/2$. We now have $2n/2^{j_0} \leq t(B')/2 \leq t(B) \leq t(B') \leq 8n/2^{j_0}$, so that $t(B) \in I_{j_0}$ and so $B \in \mathcal{B}(j_0)$.

Let $B' = B(A)$. If the event $F_B(i_0; A)$ occurs then $t(B') = t(B(A)) > 2^{i_0+10}$, with the choice of j_0 and B described in the previous paragraph we have a set $B \in \mathcal{B}(j_0)$ with $t(B) \geq t(B')/2 > 2^{i_0+9}$ and with $e(B, A) = 0$. Let j_1 be the largest integer j with $8n/2^j l > 2^{i_0+10}$, we note that as $8n/2^{j_0}l \geq t(B') > 2^{i_0+10}$ we have that $j_0 \leq j_1$. For each $j_0 \leq j_1$ and $(b_1, ..., b_{j_0'}) \in T(j_0)$ let $F_B(i_0; A; j_0; b_1, ..., b_{j_0'})$ be the event that there is a set $B \in \mathcal{B}(j_0; b_1, ..., b_{j_0'})$ with $e(B, A) = 0$. We have deduced the following inclusion of events.

$$F_B(i_0; A) \subset \bigcup_{j_0=1}^{j_1} \bigcup_{(b_1,..,b_{j_0'})\in T(j_0)} F(i_0; A; j_0; b_1, ..., b_{j_0'}) \tag{39}$$

This is a union over at most ln^l events so to prove (35) and so the Lemma we must show for every choice of $j_0 \leq j_1$ and $(b_1, .., b_{j_0'}) \in T(j_0)$ that

$$\mathbb{P}(F(i_0; A; j_0; b_1, ..., b_{j_0'})) \leq \exp(-166n/l) \tag{40}$$

Fix $j_0 \leq j_1$ and $(b_1, .., b_{j_0'}) \in T(j_0)$. Note that the probability that $e(B, A) = 0$ is the same for each $B \in \mathcal{B}(j_0; b_1, .., b_{j_0'})$. So that fixing $B \in \mathcal{B}(j_0; (b_1, .., b_{j_0'})$ we may bound the expression in (40) by,

$$|\mathcal{B}(j_0; (b_1, .., b_{j_0'})|\mathbb{P}(e(B, A) = 0) \tag{41}$$

We may bound $|\mathcal{B}(j_0; (b_1, .., b_{j_0'})|$ as we bounded $|\mathcal{A}(i_0; a_1, ..., a_{i_0'})|$ in (19) so we have that,

$$|\mathcal{B}(j_0; (b_1, .., b_{j_0'})| \leq \exp(165n/l) \tag{42}$$

It now suffices to prove

$$\mathbb{P}(e(B, A) = 0) \leq \exp(-331n/l) \tag{43}$$

We may bound $\mathbb{P}(e(B, A) = 0)$ as follows,

$$\mathbb{P}(e(B, A) = 0) = \prod_{i=1}^{i_0'} \prod_{j=1}^{j_0'} \mathbb{P}(e(B_j, A_i) = 0) = \prod_{i=1}^{i_0'} \prod_{j=1}^{j_0'} \left(1 - \frac{1}{2^{i+j}}\right)^{a_i b_j} \leq \prod_{i=1}^{i_0'} \prod_{j=1}^{j_0'} \exp\left(\frac{-a_i b_j}{2^{i+j}}\right) \tag{44}$$

alternatively this may be expressed as,

$$\mathbb{P}(e(B, A) = 0) \leq \exp\left(- \sum_{i=1}^{i_0'} \frac{a_i}{2^i} \sum_{j=1}^{j_0'} \frac{b_j}{2^j}\right) = \exp(-s(A)t(B)) \tag{45}$$

we know that $A \in \mathcal{A}(i_0)$ so that $s(A) \geq 2n/2^{i_0}l$ and $B \in \mathcal{B}(j_0)$ so that $t(B) \geq 2n/2^{j_0}l \geq 2n/2^{j_1}l > 2^{i_0+8}$. It follows that $s(A)t(B) > 2^9 n/l$, so that

$$\mathbb{P}(e(B,A)=0) \leq \exp\bigg(- \frac{512n}{l}\bigg) \tag{46}$$

which proves (43) $\square$

We now prepare to prove Lemma 2.6 which of course implies Lemma 2.2 and therefore Theorem 1.4. We have established Lemma 2.7 which states that for a fixed $i_0 \in \{1,...,l\}$ and $(a_1,...,a_{i'_0}) \in S(i_0)$ we have $\mathbb{P}(F_B(i_0;a_1,...,a_{i'_0})) \leq ln^l \exp(-n/l)$. We are now preparing to show that for each $i_0 \in \{1,...,l\}$ and $(a_1,...,a_{i'_0}) \in S_{i_0}$ we have $\mathbb{P}(F(i_0;a_1,...,a_{i'_0})) \leq 2ln^l \exp(-n/l)$. To do this we show,

$$\mathbb{P}(F(i_0;a_1,...,a_{i'_0}) \setminus F_B(i_0;a_1,...,a_{i'_0})) \leq \exp(-n/l) \tag{47}$$

In other words we show that if no set $A \in \mathcal{A}(i_0;a_1,...,a_{i'_0})$ has $t(B(A)) > 2^{i_0+10}$ then with high probability no set $A \in \mathcal{A}(i_0;a_1,...,a_{i'_0})$ has $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > Kn/l$. Let us note that

$$s(A \cup \bigcup_{i_0+1}^{l/2} U_i) \leq s(A) + \alpha \sum_{i=i_0+1}^{l/2} \frac{n}{l2^i} \leq \frac{8n}{2^{i_0}l} + \frac{\alpha n}{l2^{i_0}} \leq \frac{10n}{2^{i_0}l} \tag{48}$$

Let a'_i be defined by $a'_i = |A_i| = a_i$ for $i \leq i_0$ and $a'_i = |U_i|$ for $i > i_0$, so that $\sum_i a'_i/2^i = s(A \cup \bigcup_{i_0+1}^{l/2} U_i) \leq 10n/2^{i_0}l$.

Let us study the distribution of $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$. For each pair uv with $u \in U_i$ and $v \in V_j$ let $p_{uv} = \mathbb{P}(uv \in E(G)|E(V,U))$ we have $p_{uv} \leq 1/2^{i+j-1}$ (see remark after the definition of G). We define the random variable X_{uv} to be 1 if $uv \in E(G)$ and 0 otherwise. Then of course $\{X_{uv} : u \in U, v \in V\}$ is an independent family and $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$ may be expressed as

$$e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) = \sum_{u \in A \cup \bigcup_{i_0+1}^{l/2} U_i} \sum_{v \in B} X_{uv} \tag{49}$$

An upper bound on the expectation of $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$ is now given by

$$\sum_{i=1}^{l/2} \sum_{j=1}^{l/2} \frac{a'_i b_j}{2^{i+j-1}} = 2s(A \cup \bigcup_{i_0+1}^{l/2} U_i)t(B) \leq 2\frac{10n}{2^{i_0}l}2^{i_0+10} \leq \frac{21000n}{l} \tag{50}$$

We wish to show that the probability that $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$ is larger than Kn/l is very small. We require a result about large deviations, see for example [2].

Lemma 2.9. *[(Chernoff bound for the sum of Poisson trials)] Let $\beta > 0$ and let $S_N = X_1 + ... + X_N$ be a sum of independent random variables X_i taking values in $\{0,1\}$, with $\mathbb{P}(X_i = 1) = p_i$ for each $i = 1,...,N$ so that the expectation of S_N is $\mu = \sum_i p_i$. Then*

$$\mathbb{P}(S_N \geq (1+\beta)\mu) \leq \exp(\frac{-\beta^2\mu}{3})$$

Furthermore, for any constant $\lambda \geq \mu$ we have

$$\mathbb{P}(S_N \geq (1+\beta)\lambda) \leq \exp(\frac{-\beta^2\lambda}{3})$$

In the other direction we have,

$$\mathbb{P}(S_N \leq (1-\beta)\mu) \leq \exp(\frac{-\beta^2\mu}{2})$$

Proof. The first and third statements are well known results in probability theory, proofs can be found in [2]. The second statement follows from the first by a monotonicity argument. Define $X_{N+1}, ..., X_{N'}$ such that $\sum_{i=1}^{N'} p_i = \lambda$ the the result follows from applying the first result to $S_{N'}$ and noting $S_N \leq S_{N'}$. $\square$

We can express $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$ as a sum of independent $\{0,1\}$-valued random variables. The above Lemma allows us to show that with high probability this value does not exceed twice it's expectation, this is exactly what is required to complete the proof of Lemma 2.6. Recall that we are currently proving the result for even n and with $K = 50000$.

Proof of Lemma 2.6. By (16) and Lemma 2.7 it suffices to prove that,

$$\mathbb{P}(F(i_0; a_1, ..., a_{i_0'}) \setminus F_B(i_0; a_1, ..., a_{i_0'})) \leq \exp\left(\frac{-n}{l}\right) \tag{51}$$

The event $F(i_0; a_1, ..., a_{i_0'}) \setminus F_B(i_0; a_1, ..., a_{i_0'})$ is the union over $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ of the events $F(i_0; A) \setminus F_B(i_0; A)$. We have from Lemma 2.8 that $|\mathcal{A}(i_0; a_1, ..., a_{i_0'})| \leq \exp(165n/l)$ so that it suffices to prove for each $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ that

$$\mathbb{P}(F(i_0; A) \setminus F_B(i_0; A)) \leq \exp\left(\frac{-166n}{l}\right) \tag{52}$$

Fix $A \in \mathcal{A}(i_0; a_1, ..., a_{i_0'})$ we must show that the probability that $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > Kn/l$ given that $t(B(A)) \leq 2^{i_0+10}$ is at most $\exp(-166n/l)$. We have seen that the expectation of $e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A))$, which we denote μ, is at most $21000n/l$. We note that $K/21000 > 2$, so setting $\lambda = 21000n/l$ and applying Lemma 2.9 we obtain

$$\mathbb{P}\left(e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > \frac{Kn}{l}\right) \leq \mathbb{P}\left(e(A \cup \bigcup_{i_0+1}^{l/2} U_i, B(A)) > 2\lambda\right) \leq \exp\left(\frac{-\lambda}{3}\right) = \exp\left(\frac{-7000n}{l}\right)$$

$\square$

3 Proof of Theorem 1.5

We prove Theorem 1.5 using as our example the oriented graph G defined in Section 2. We already know that with high probability G has no isolated vertices, and so it suffices to show

that if K_ν is a sufficiently large constant then with high probability G contains no pair of subsets $A, B \subset V$ with

$$e(A,B) > \frac{K_\nu n}{l} \qquad \text{and} \qquad e(B,A) < \nu e(A,B) \tag{53}$$

We will see that this follows from the following Lemma,

Lemma 3.1. *Given $\eta \in (0,1)$ there is a constant L_η such that*

$$\mathbb{P}\Bigg(e(A,B) \geq \frac{L_\eta n}{l} \quad \text{for some} \quad A \subset U, B \subset V \quad \text{with} \quad e(B,A) < \eta e(A,B)\Bigg) \leq \exp(-n/l)$$

By symmetry we have that the result of Lemma 3.1 holds also for $A \subset V$ and $B \subset U$. We know prove (assuming Lemma 3.1) that with high probability there is no pair $A, B \subset V$ with

$$e(A,B) > \frac{K_{nu} n}{l} \qquad \text{and} \qquad e(B,A) < \nu e(A,B) \tag{54}$$

we do this by showing that this event is contained in $F^\nu_{UV} \cup F^\nu_{VU} \cup F^\eta_{UV} \cup F^\eta_{VU}$ where F^η_{UV} is the event that there are subsets $A \subset U, B \subset V$ with $e(A,B) \geq \frac{L_\eta n}{l}$ and $e(B,A) < \eta e(A,B)$ and $F^\nu_{UV}, F^\nu_{VU}, F^\eta_{VU}$ are defined similarly. The Lemma shows that the union of these events has probability at most $4\exp(-n/l)$, and so we are done.

Proof of Theorem 1.5 assuming Lemma 3.1. Set $\eta = (1+\nu)/2$ and set

$$K_\nu = \max\{L_\eta + L_\nu, (1 + \frac{1}{\eta - \nu})L_\nu\} \tag{55}$$

We show that if none of the events $F^\nu_{UV}, F^\nu_{VU}, F^\eta_{UV}, F^\eta_{VU}$ occur then no pair of subsets $A, B \subset V$ can have $e(A,B) > K_\nu n/l$ and $e(B,A) < \nu e(A,B)$, our proof is by contradiction. Suppose some pair of subsets $A, B \subset V$ have $e(A,B) > K_\nu n/l$ and $e(B,A) < \nu e(A,B)$, then we define

$$A_U = A \cap U,\ A_V = A \cap V,\ B_U = B \cap U, \text{ and } B_V = B \cap V \tag{56}$$

since G is bipartite we have that $e(A,B) = e(A_U, B_V) + e(A_V, B_U)$ and $e(B,A) = e(B_V, A_U) + e(B_U, A_V)$, it follows that $e(B,A)/e(A,B)$ is a weighted average of $e(B_V, A_U)/e(A_U, B_V)$ and $e(B_U, A_V)/e(A_V, B_U)$. So that one of these quantities must be less than ν, so that with loss of generality we may assume,

$$e(B_V, A_U) < \nu e(A_U, B_V) \tag{57}$$

Since we are assuming F^ν_{UV} does not occur we have $e(A_U, B_V) \leq L_\nu n/l$. We now consider the ratio $e(B_U, A_V)/e(A_V, B_U)$ if it is less than η then since F^η_{VU} does not occur we have $e(A_V, B_U) \leq L_\eta n/l$, whence

$$e(A,B) = e(A_U, B_V) + e(A_V + B_U) \leq (L_\nu + L_\eta)\frac{n}{l} \leq K_\nu \frac{n}{l} \tag{58}$$

which is a contradiction. So we may assume that $e(B_U, A_V)/e(A_V, B_U) \geq \eta$, whence

$$e(A_V, B_U) \leq \frac{e(B_U, A_V)}{\eta} \leq \frac{e(B,A)}{\eta} < \frac{\nu e(A,B)}{\eta} \tag{59}$$

while $e(A,B) = e(A_U, B_V) + e(A_V, B_U) \le e(A_V, B_U) + L_\nu n/l$ so that,

$$(1 - \frac{\nu}{\eta}) e(A_V, B_U) \le \frac{\nu L_\nu n}{\eta l} \tag{60}$$

multiplying by η and then dividing by $\eta - \nu$ we obtain,

$$e(A_V, B_U) \le \frac{\nu L_\nu n}{(\eta - \nu) l} \le \frac{L_\nu n}{(\eta - \nu) l} \tag{61}$$

so that

$$e(A,B) = e(A_V, B_U) + e(A_U, B_V) \le \frac{L_\nu n}{(\eta - \nu) l} + \frac{L_\nu n}{l} \le \frac{K_\nu n}{l} \tag{62}$$

a contradiction. □

We now prove Lemma 3.1. Fix $\eta \in (0,1)$. We define,

$$\delta = 1 - \eta \qquad k = \lceil \log(64/\delta) \rceil \qquad \text{and} \qquad L_\eta = \frac{9216(350 + 2k)}{\delta^2} \tag{63}$$

Since η is fixed we write L for L_η. We shall use many of the same notions and indeed notations as were used in Section 2, including $A_i, a_i, s(A), B_j, b_j, t(B)$.

In the previous Section we wished to prove that with high probability $e(A, B(A)) \le Kn/l$ for all $A \subset U$. Probabilistic methods allowed us to do this, but only after we had greatly reduced the number of 'bad' events we considered. For that reason we considered for a set $A' \subset U$, a restriction A of A' given by $A = A' \cap \bigcup_{i=1}^{i_0} U_i$ and the extension of A' to $A \cup \bigcup_{i > i_0} U_i$. This approximation was rather crude, it required only that the 'size' of A and the 'size' of $A \cup \bigcup_{i > i_0} U_i$ were of the same order. To prove Lemma 3.1 we require a more precise approximation.

We now define new versions of the collections $\mathcal{A}(i_0)$ and $\mathcal{B}(j_0)$ which will allow our approximations to be more precise. For $i_0 \in \{1, ..., l\}$ we define the interval,

$$I_{i_0}^k = \left[\frac{2^k n}{2^{i_0} l}, \frac{2^{k+2} n}{2^{i_0} l} \right) \tag{64}$$

and the collection

$$\mathcal{A}^k(i_0) = \{ A \subset \bigcup_{i=1}^{i_0'} U_i : s(A) \in I_{i_0}^k \} \tag{65}$$

We also define a partition of $\mathcal{A}^k(i_0)$. For a sequence $a_1, ..., a_{i_0'}$ we let $s(a_1, ..., a_{i_0'}) = \sum_i a_i/2^i$ and write $S^k(i_0)$ for the set of sequences $(a_1, ..., a_{i_0'})$ with $a_i \le |U_i|$ for each i and with $s(a_1, ..., a_{i_0'}) \in I_{i_0}^k$. For each $(a_1, ..., a_{i_0'}) \in S^k(i_0)$ we define,

$$\mathcal{A}^k(i_0; a_1, ..., a_{i_0'}) = \{ A \in \bigcup_{i=1}^{i_0'} U_i : |A_i| = a_i \text{ for all } i \} \tag{66}$$

Of course we may now partition $\mathcal{A}^k(i_0)$ as follows

$$\mathcal{A}^k(i_0) = \bigcup_{(a_1, ..., a_{i_0'}) \in S^k(i_0)} \mathcal{A}^k(i_0; a_1, ..., a_{i_0}) \tag{67}$$

We similarly define for $j_0 \in \{1, ..., l\}$ the interval $I_{j_0}^k = [\frac{2^k n}{2^{j_0} l}, \frac{2^{k+2} n}{2^{j_0} l})$, the set $S^k(j_0)$, the collection

$$\mathcal{B}^k(j_0) = \{B \subset \bigcup_{j=1}^{j_0'} V_j : t(B) \in I_{j_0}^k\} \tag{68}$$

and for each sequence $(b_1, ..., b_{j_0'}) \in S^k(j_0)$,

$$\mathcal{B}^k(j_0; b_1, ..., b_{j_0'}) = \{B \in \bigcup_{j=1}^{j_0'} V_j : |B_j| = b_j \text{ for all } j\} \tag{69}$$

As in the previous section it is important to bound the size of the collections $\mathcal{A}^k(i_0; a_1, ..., a_{i_0'})$.

Lemma 3.2. *Let $i_0 \in \{1, ..., l\}$ and $(a_1, ..., a_{i_0'}) \in S^k(i_0)$ then $|\mathcal{A}^k(i_0; a_1, ..., a_{i_0'})| \le \exp((165 + k)n/l)$*

Proof. For every sequence $a_1, ..., a_{i_0'}$ with $s(a_1, ..., a_{i_0'}) \le 4n/2^{i_0} l$, let us define,

$$\mathcal{A}^*(i_0; a_1, ..., a_{i_0'}) = \{A \subset \bigcup_{i=1}^{i_0'} U_i : |A_i| = a_i \text{ for all } i\} \tag{70}$$

. We claim that for each sequence $a_1, ..., a_{i_0'}$ with $s(a_1, ..., a_{i_0'}) \le 4n/2^{i_0} l$ we have that $|\mathcal{A}^*(i_0; a_1, ..., a_{i_0'})| \le \exp(165n/l)$, for a proof simply proceed as in the proof of Lemma 2.8, in fact in that proof we use only that $s(a_1, ..., a_{i_0'}) \le 8n/2^{i_0} l$. Now for a given set $A \in \mathcal{A}^k(i_0; a_1, ..., a_{i_0'})$ the restriction A^- of A to $\bigcup_{i=1}^{i_0-k} U_i$ has $s(A^-) \le s(A) \le 2^{k+2} n/2^{i_0} l \le 4n/2^{i_0-k} l$ and so $A^- \in \mathcal{A}^*(i_0; a_1, ..., a_{i_0-k})$. For each such A^- there are at most $2^{\alpha k n/l}$ extensions of A^- to a set contained in $\bigcup_{i=1}^{i_0'}$ and so at most $2^{\alpha k n/l}$ extensions of A^- to a set $A \in \mathcal{A}^k(i_0; a_1, ..., a_{i_0'})$, it follows that,

$$|\mathcal{A}^k(i_0; a_1, ..., a_{i_0'})| \le |\mathcal{A}^*(i_0; a_1, ..., a_{i_0'})| 2^{\alpha k n/l} \le \exp\left(\frac{(165 + k)n}{l}\right) \tag{71}$$

□

Similarly for $j_0 \in \{1, ..., l\}$ and $(b_1, ..., b_{j_0'}) \in S^k(j_0)$ we have

$$|\mathcal{B}^k(j_0; b_1, ..., b_{j_0'})| \le \exp\left(\frac{(165 + k)n}{l}\right) \tag{72}$$

it is even possible to put bounds of this type on the cardinality of,

$$\mathcal{A}^k = \bigcup_{i_0=1}^{l} \bigcup_{(a_1, ..., a_{i_0'}) \in S(i_0)} \mathcal{A}^k(i_0; a_1, ..., a_{i_0'}) \tag{73}$$

indeed this is a union over l choices of i_0 and for each i_0 there are at most $(\alpha n/l)^{i_0} \leq n^l$ choices of a sequence $(a_1, ..., a_{i_0'}) \in S(i_0)$, so that from Lemma 3.2 we have,

$$|\mathcal{A}^k| \leq ln^l \exp\left(\frac{(165+k)n}{l}\right) \leq \exp\left(\frac{(170+k)n}{l}\right) \tag{74}$$

the last inequality is somewhat arbitrary, but certainly holds for all reasonably large n (and for small n all the results are trivial). Defining,

$$\mathcal{B}^k = \bigcup_{j_0=1}^{l} \bigcup_{(b_1,...,b_{j_0'}) \in S(i_0)} \mathcal{B}^k(j_0; b_1, ..., b_{j_0'}) \tag{75}$$

we have similarly that

$$|\mathcal{B}^k| \leq \exp\left(\frac{(170+k)n}{l}\right) \tag{76}$$

As in the previous Section we shall consider sets $A \in \mathcal{A}^k(i_0; a_1, ..., a_{i_0'})$ and $B \in \mathcal{B}^k(j_0; b_1, ..., b_{j_0'})$ together with their extensions $A \cup \bigcup_{i>i_0} U_i$ and $B \cup \bigcup_{j>j_0'} V_j$. We shall show that with high probability the number of edges between sets $A \in \mathcal{A}^k(i_0; a_1, ..., a_{i_0'})$ and $B \in \mathcal{B}^k(j_0; b_1, ..., b_{j_0'})$ and the orientations of these edges are as we would expect. We then simply need to ensure that the collections of edges $E(A, \bigcup_{j>j_0} V_j)$, $E(\bigcup_{i>i_0} U_i, B)$ and $E(\bigcup_{i>i_0} U_i, \bigcup_{j>j_0} V_j)$ are (with high probability) sufficiently small that they do not interfere with the situation observed between A and B.

We define some 'bad' events, show that with high probability none of them occur and finally show that if none of these events occur then there is no pair $A \subset U$ and $B \subset V$ with $e(A,B) \geq Ln/l$ and $e(B,A) < \eta e(A,B)$. For $i_0 \in \{1, ..., l\}$ and $j_0 \in \{1, ..., l\}$ we define

$$\mu(i_0, j_0) = 2\left(\frac{2^k n}{2^{i_0} l}\right)\left(\frac{2^k n}{2^{j_0} l}\right) \tag{77}$$

Let $\bar{e}(A,B)$ be the number of edges between A and B in the underlying graph, ie. $\bar{e}(A,B) = e(A,B) + e(B,A)$. Let us also define the quantity $\tilde{e}(i_0, A; j_0, B)$ to be the number of edges between $A \cup \bigcup_{i>i_0} U_i$ and $B \cup \bigcup_{j>j_0} V_j$ which are not in fact between A and B, alternatively $\tilde{e}(i_0, A; j_0, B) = |\widetilde{E}(i_0, A; j_0, B)|$ where,

$$\widetilde{E}(i_0, A; j_0, B) = \bar{E}(A, \bigcup_{j>j_0} V_j) \cup \bar{E}(\bigcup_{i>i_0} U_i, B) \cup \bar{E}(\bigcup_{i>i_0} U_i, \bigcup_{j>j_0} V_j) \tag{78}$$

Finally, let let $\xi = (1+\eta)/2$. We define the four 'bad' events F_I, F_{II}, F_{III} and F_{IV} as follows,

- F_I : There is a pair i_0, j_0 with $\mu(i_0, j_0) < \frac{Ln}{128l}$ and sets $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j) \geq \frac{Ln}{l}$
- F_{II} : There is a pair i_0, j_0 with $\mu(i_0, j_0) \geq \frac{Ln}{128l}$ and sets $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\bar{e}(A,B) < s(A)t(B)$
- F_{III} : There is a pair i_0, j_0 with $\mu(i_0, j_0) \geq \frac{Ln}{128l}$ and sets $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\bar{e}(A,B) \geq s(A)t(B)$ and $e(B,A) < \xi e(A,B)$
- F_{IV} : There is a pair i_0, j_0 with $\mu(i_0, j_0) \geq \frac{Ln}{128l}$ and sets $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\bar{e}(A,B) \geq s(A)t(B)$ and $\tilde{e}(i_0, A; j_0, B) > \delta \bar{e}(A,B)/4$

Lemma 3.3. $\mathbb{P}(F_I) \leq \exp(-(10)n/l)$

Proof. By 74 and 75 we have that the event F_I is the union over at most $\exp((340+2k)n/l)$ choices A, B, so that it suffices to show that for each pair $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\mu(i_0, j_0) < \frac{Ln}{128l}$ we have,

$$\mathbb{P}\bigg(\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j) \geq \frac{Ln}{l}\bigg) \leq \exp\left(\frac{(350+2k)n}{l}\right) \tag{79}$$

For such a pair A, B, we note that the quantity $\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j)$ is the sum of independent $\{0,1\}$-valued random variables X_{uv}. Define the sequence $a'_i = |A_i|$ for $i = 1, ..., i'_0$ and $a'_i = |U_i|$ for $i > i_0$ and similarly $b'_j = |B_j|$ for $j = 1, ..., j'_0$ and $b'_j = |V_j|$ for $j > j_0$. The expectation of $\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j)$ is,

$$\sum_{i=1}^{l/2} \sum_{j=1}^{l/2} \frac{a'_i b'_j}{2^{i+j-1}} = 2s(a'_1, ..., a'_{l/2})t(b'_1, ..., b'_{l/2}) \tag{80}$$

however we have that,

$$s(a'_1, ..., a'_{l/2}) = \sum_{i=1}^{i'_0} \frac{a_i}{2^i} + \sum_{i>i_0} \frac{\alpha n}{l2^i} \leq s(A) + \frac{2n}{2^{i_0}l} \tag{81}$$

similarly $t(b'_1, ..., b'_{l/2}) \leq t(B) + 2n/2^{j_0}l$. Of course we also have that $s(A) \leq 2^{k+2}n/2^{i_0}l$ and $t(B) \leq 2^{k+2}n/2^{j_0}l$. So that we may bound above the expectation of $\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j)$ by

$$2\bigg(s(A) + \frac{2n}{2^{i_0}l}\bigg)\bigg(t(B) + \frac{2n}{2^{j_0}l}\bigg) \leq 2\bigg(\frac{2^{k+3}n}{2^{i_0}l}\bigg)\bigg(\frac{2^{k+3}n}{2^{j_0}l}\bigg) = 64\mu(i_0, j_0) \tag{82}$$

by our assumption that $\mu(i_0, j_0) < Ln/128l$ we have that

$$\mathbb{E}(\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j)) \leq \frac{Ln}{2l} \tag{83}$$

Applying Lemma 2.9 with $\lambda = Ln/2l$ we obtain

$$\mathbb{P}(F^k(A, B)) \leq \exp\left(\frac{-\lambda}{3}\right) = \exp\left(\frac{Ln}{6l}\right) \tag{84}$$

and so we are done because $L \geq 6(350+2k)$. $\square$

Lemma 3.4. $\mathbb{P}(F_{II}) \leq \exp(-10n/l)$

Proof. By arguing as in the previous proof, it suffices to show for a fixed $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\mu(i_0, j_0) \geq Ln/128l$ that

$$\mathbb{P}(\bar{e}(A, B) < s(A)t(B)) \leq \exp\left(\frac{(350+2k)n}{l}\right) \tag{85}$$

Note that $\bar{e}(A,B)$ is a sum of independent $\{0,1\}$-valued random variables and that

$$\mathbb{E}(\bar{e}(A,B)) = \sum_{i=1}^{i_0'}\sum_{j=1}^{j_0'} \frac{a_i b_j}{2^{i+j-1}} = 2s(A)t(B) \tag{86}$$

applying Lemma 2.9 we obtain that

$$\mathbb{P}\big(\bar{e}(A,B) < s(A)t(B)\big) \leq \exp\left(\frac{-s(A)t(B)}{8}\right) \tag{87}$$

in fact this is a very good bound on the probability because $s(A) \geq 2^k n/2^{i_0} l$ and $t(B) \geq 2^k n/2^{j_0} l$, and so,

$$\mathbb{P}\big(\bar{e}(A,B) < s(A)t(B)\big) \leq \exp\left(\frac{-1}{8}\cdot\frac{2^k n}{2^{i_0} l}\cdot\frac{2^k n}{2^{j_0} l}\right) = \exp\left(\frac{-\mu(i_0,j_0)}{8}\right) \leq \exp\left(\frac{-Ln}{1024l}\right) \tag{88}$$

we are now done as $L \geq 1024(350+2k)$. □

Lemma 3.5. $\mathbb{P}(F_{III}) \leq \exp(-10n/l)$

Proof. It suffices to show for a fixed $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\mu(i_0,j_0) \geq Ln/128l$ and with $\bar{e}(A,B) \geq s(A)t(B)$, that

$$\mathbb{P}(e(B,A) < \xi e(A,B)) \leq \exp\left(\frac{(350+2k)n}{l}\right) \tag{89}$$

Recalling that $\bar{e}(A,B) = e(A,B) + e(B,A)$ and re-arranging the inequality $e(B,A) < \xi e(A,B)$ we see that it is equivalent to the statement that $e(B,A) < \xi\bar{e}(A,B)/(1+\xi)$. Define also $\gamma = 1 - 2\xi/(1-\xi)$. As each underlying edge between A and B is oriented with probability $1/2$ of going in each direction we have that $e(B,A)$ can be expressed as a sum of independent $\{0,1\}$-valued random variables and $\mathbb{E}(e(B,A)) = \bar{e}(A,B)/2$ so that we may apply Lemma 2.9 and deduce that,

$$\mathbb{P}\Big(e(B,A) < \xi e(A,B)\Big) = \mathbb{P}\Big(e(B,A) < \frac{\xi}{1-\xi}\bar{e}(A,B)\Big) = \mathbb{P}\left(e(B,A) < (1-\gamma)\frac{\bar{e}(A,B)}{2}\right) \leq \exp\left(\frac{-\gamma^2\bar{e}(A,B)}{4}\right) \tag{90}$$

from our assumptions we have that $\bar{e}(A,B) \geq s(A)t(B) \geq \mu(i_0,j_0)/2 \geq Ln/256l$ and a little calculation shows that $\gamma \geq \delta/3$ so that,

$$\mathbb{P}(e(B,A) < \xi e(A,B)) \leq \exp\left(\frac{-\delta^2 Ln}{9216l}\right) \tag{91}$$

and so we are done as we may assume that $L = 9216(350+2k)/\delta^2$ □

Lemma 3.6. $\mathbb{P}(F_{IV}) \leq \exp(-10n/l)$

Proof. It suffices to show for a fixed $A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0)$ with $\mu(i_0, j_0) \geq Ln/128l$ and with $\bar{e}(A,B) \geq s(A)t(B)$, that

$$\mathbb{P}(\tilde{e}(i_0, A; j_0, B) > \delta \bar{e}(A,B)/4) \leq \exp\left(\frac{(350+2k)n}{l}\right) \tag{92}$$

We may express the quantity $\tilde{e}(i_0, A; j_0, B)$ as,

$$\tilde{e}(i_0, A; j_0, B) = \bar{e}(A, \bigcup_{j>j_0} V_j) + \bar{e}(\bigcup_{i>i_0} U_i, B) + \bar{e}(\bigcup_{i>i_0} U_i, \bigcup_{j>j_0} V_j) \tag{93}$$

Note that $\tilde{e}(i_0, A; j_0, B)$ may be expressed as a sum of independent $\{0,1\}$-valued random variables and it's expectation is given by

$$\mathbb{E}(\tilde{e}(i_0, A; j_0, B)) = \mathbb{E}(\bar{e}(A, \bigcup_{j>j_0} V_j)) + \mathbb{E}(\bar{e}(\bigcup_{i>i_0} U_i, B)) + \mathbb{E}(\bar{e}(\bigcup_{i>i_0} U_i, \bigcup_{j>j_0} V_j)) \tag{94}$$

so that using the sequences a'_i, b'_j defined in the proof of Lemma 3.3 we have,

$$\mathbb{E}(\tilde{e}(i_0, A; j_0, B)) = \sum_{i=1}^{i_0} \sum_{j>j_0} \frac{a'_i b'_j}{2^{i+j-1}} + \sum_{i>i_0} \sum_{j=1}^{j_0} \frac{a'_i b'_j}{2^{i+j-1}} + \sum_{i>i_0} \sum_{j>j_0} \frac{a'_i b'_j}{2^{i+j-1}} \tag{95}$$

These expressions are easily bounded as for $i > i_0$ and $j > j_0$ we have $a'_i = |U_i| \leq \alpha n/l$ and $b'_j \leq \alpha n/l$ so that,

$$\mathbb{E}(\tilde{e}(i_0, A; j_0, B)) \leq \frac{\alpha n\, s(A)}{2^{j_0} l} + \frac{\alpha n\, t(B)}{2^{i_0} l} + \frac{\alpha^2 n^2}{2^{i_0+j_0} l^2} \tag{96}$$

Since $s(A) \geq 2^k n/2^{i_0} l$ and $t(B) \geq 2^k n/2^{j_0} l$ it follows that,

$$\mathbb{E}(\tilde{e}(i_0, A; j_0, B)) \leq \frac{s(A)t(B)}{2^k}\left(\alpha + \alpha + \frac{\alpha^2}{2^k}\right) \leq \frac{s(A)t(B)}{2^{k-2}} \tag{97}$$

recalling that $2^k \geq 64/\delta$ we have

$$\mathbb{E}(\tilde{e}(i_0, A; j_0, B)) \leq \frac{\delta s(A)t(B)}{16} \tag{98}$$

Applying Lemma 2.9 with $\lambda = \delta s(A)t(B)/16$ we deduce that,

$$\mathbb{P}\left(\tilde{e}(i_0, A; j_0, B) > \frac{\delta \bar{e}(A,B)}{4}\right) \leq \mathbb{P}\left(\tilde{e}(i_0, A; j_0, B) > \frac{\delta s(A)t(B)}{4}\right) \leq \exp\left(\frac{-s(A)t(B)}{8}\right) \tag{99}$$

from our assumptions we have that $s(A)t(B) \geq \mu(i_0, j_0)/2 \geq Ln/256l$ and so we are done as $L \geq 2048(350+2k)$. □

Writing F_{bad} for the union of all the 'bad' events $F_I, F_{II}, F_{III}, F_{IV}$ we have $\mathbb{P}(F_{bad}) \leq \exp(-9n/l)$. We write F for the event that there is a pair $A' \subset U$ and $B' \subset V$ with $e(A', B') \geq$

Ln/l and $e(B', A') < \eta e(A', B')$, thus to prove Lemma 3.1 we must show $\mathbb{P}(F) \le \exp(-n/l)$. To do this it clearly suffices to show $F \subset F_{bad}$. We do this by supposing that none of the 'bad' events occurs and showing that this implies that F does not occur. Therefore we assume for i_0, j_0 with $\mu(i_0, j_0) < Ln/128l$ that,

$$\bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j) < Ln/l \qquad \text{for all} \quad A \in \mathcal{A}^k(i_0), B \in \mathcal{B}^k(j_0) \tag{100}$$

While for i_0, j_0 with $\mu(i_0, j_0) \ge Ln/128l$ we have for all pairs $A \in \mathcal{A}^k(i_0)$ and $B \in \mathcal{B}^k(j_0)$ that,

$$\bar{e}(A, B) \ge s(A)t(B) \quad , \quad e(B, A) \ge \xi e(A, B) \qquad \text{and} \qquad \tilde{e}(i_0, A; j_0, B) \le \frac{\delta \bar{e}(A, B)}{4} \tag{101}$$

Lemma 3.7. *If (100) and (101) hold then no pair $A' \subset U$ and $B' \subset V$ can have $e(A', B') \ge Ln/l$ and $e(B', A') < \eta e(A', B')$*

Proof. Given $A' \subset U$ and $B' \subset V$ let i_0 be chosen such that $s(A') \in [2^{k+1}n/l2^{i_0}, 2^{k+2}n/l2^{i_0})$ and let j_0 be chosen such that $t(B') \in [2^{k+1}n/l2^{j_0}, 2^{k+2}n/l2^{j_0})$. Let $A = A' \cap \bigcup_{i=1}^{i_0'} U_i$ and $B = B' \cap \bigcup_{j=1}^{j_0'} V_j$ it follows that $s(A) \le s(A') \le 2^{k+2}n/2^{i_0}l$ and $s(A) \ge s(A') - \alpha n/2^{i_0}l \ge 2^k n/2^{i_0}l$ so that $A \in \mathcal{A}^k(i_0)$, similarly $B \in \mathcal{B}^k(j_0)$. Let us also note that,

$$\bar{E}(A, B) \subset \bar{E}(A', B') \subset \bar{E}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j) \tag{102}$$

If $\mu(i_0, j_0) < Ln/128l$ then from (102) and (100) we have,

$$\bar{e}(A', B') \le \bar{e}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j) < Ln/l \tag{103}$$

and so the conditions do not hold for such pairs A', B'. If $\mu(i_0, j_0) \ge Ln/128l$ then we have from (101) that $e(B, A) \ge \xi e(A, B)$, let us write $\tilde{e}$ for the number of edges between A' and B' which are not in $\bar{E}(A, B)$, then since $\bar{E}(A', B') \subset \bar{E}(A \cup \bigcup_{i>i_0} U_i, B \cup \bigcup_{j>j_0} V_j)$ we have $\tilde{e} \le \tilde{e}(i_0, A; j_0, B) \le \delta \bar{e}(A, B)/4$. Let us now consider the orientation of edges. We have $e(A', B') \le e(A, B) + \tilde{e} \le e(A, B) + \delta \bar{e}(A, B)/4$. Suppose that A', B' contradict the Lemma, then $e(B', A') < \eta e(A', B')$, from which it follows that,

$$\frac{e(B, A)}{\eta} \le \frac{e(B', A')}{\eta} < e(A', B') \le e(A, B) + \frac{\delta \bar{e}(A, B)}{4} \tag{104}$$

recalling that $\bar{e}(A, B) = e(A, B) + e(B, A)$ this implies $e(B, A)/\eta < (1 + \delta/4)e(A, B) + \delta e(B, A)/4$ and so,

$$e(B, A) < \frac{(1 + \frac{\delta}{4})}{(\frac{1}{\eta} - \frac{\delta}{4})} e(A, B) \tag{105}$$

a simple calculation shows that this fraction is less than ξ, a contradiction of (101). $\square$

References

[1] O. Amini and S. Griffith and F. Huc, Cycles in oriented graphs, preprint.

[2] M. Mitzenmach and E. Upfal, Proability and Computing. CUP.